

# Crime Forecasting Using Spatio-temporal Pattern with Ensemble Learning

Chung-Hsien Yu<sup>1</sup>, Wei Ding<sup>1</sup>, Ping Chen<sup>1</sup>, and Melissa Morabito<sup>2</sup>

<sup>1</sup> University of Massachusetts Boston,  
100 Morrissey Blvd., Boston, MA 02125, USA  
{csyu,ding}@cs.umb.edu, Ping.Chen@umb.edu  
<sup>2</sup> University of Massachusetts Lowell,  
One University Avenue, Lowell, MA 01854, USA  
Melissa\_Morabito@uml.edu

**Abstract.** Crime forecasting is notoriously difficult. A crime incident is a multi-dimensional complex phenomenon that is closely associated with temporal, spatial, societal, and ecological factors. In an attempt to utilize all these factors in crime pattern formulation, we propose a new feature construction and feature selection framework for crime forecasting. A new concept of multi-dimensional feature denoted as spatio-temporal pattern, is constructed from local crime cluster distributions in different time periods at different granularity levels. We design and develop the Cluster-Confidence-Rate-Boosting (CCRBoost) algorithm to efficiently select relevant local spatio-temporal patterns to construct a global crime pattern from a training set. This global crime pattern is then used for future crime prediction. Using data from January 2006 to December 2009 from a police department in a northeastern city in the US, we evaluate the proposed framework on residential burglary prediction. The results show that the proposed CCRBoost algorithm has achieved about 80% on accuracy in predicting residential burglary using the grid cell of 800-meter by 800-meter in size as one single location.

**Keywords:** Spatio-temporal Pattern, Crime Forecasting, Ensemble Learning, Boosting.

## 1 Introduction

Crime forecasting is notoriously difficult. A crime incident is a multi-dimensional complex phenomenon that is closely associated with temporal, spatial, societal, and ecological factors. In an attempt to utilize all these factors in crime pattern formulation, we propose a new feature construction and feature selection framework for crime forecasting. A new concept of multi-dimensional feature denoted as spatio-temporal pattern, is constructed from local crime cluster distributions in different time periods at different granularity levels.

Crime distributions are of different sizes and shapes with respect to spatial space over time. We use clustering to find local crime distributions in different time periods. The spatial-temporal patterns then are induced from each

crime distribution through classification. Each spatio-temporal pattern uses related crime incidences as indicators to represent a local crime pattern at certain clustered locations during a certain time period. However, these locally learned patterns could be redundant or overlapping at global level.

We design and develop the Cluster-Confidence-Rate-Boosting (CCRBoost) algorithm to efficiently select relevant local spatio-temporal patterns to construct a global crime pattern from a training set. The main idea of this approach is to iteratively pick a set of local patterns which give the least classification error at each boosting round. Each set of local patterns is referred as an ensemble spatio-temporal pattern and is assigned a score (Called confidence-rate in our approach). At the end of boosting, a global pattern is constructed from these ensemble patterns. This global pattern is capable of predicting crime by scaling the total score of an input, a collection of crime indicators, evaluated on each crafted ensemble patterns. The proposed algorithm is depicted in Figure 1.

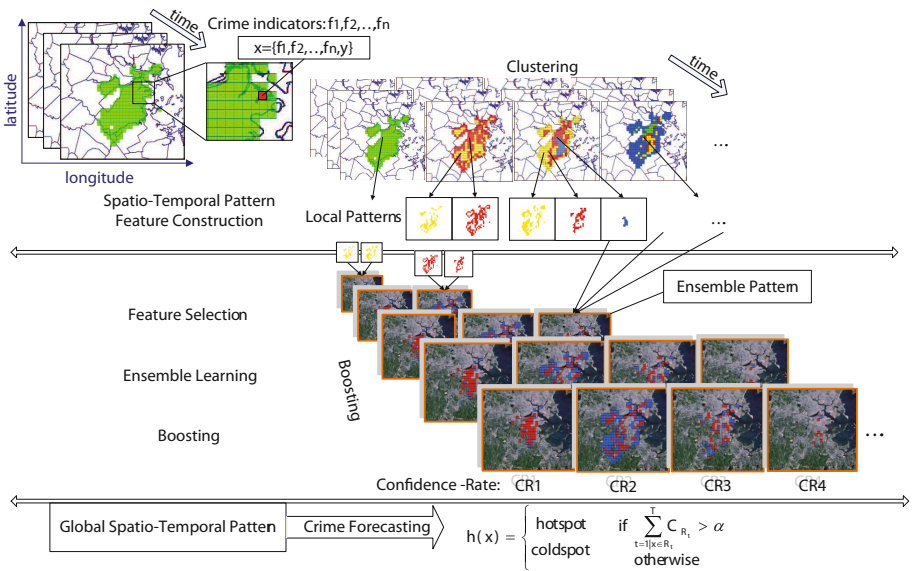


Fig. 1. The flowchart of the proposed CCRBoost algorithm (Better viewed in color)

In our real-world case study, we collaborated with the police department of a northeastern city in the US to collect 4-year historical crime data, from January 2006 to December 2009. These data are used to evaluate the proposed framework on residential burglary prediction. This city is 90 square miles in size and more than 600 thousands in population. The results show that the proposed CCRBoost algorithm has achieved about 80% on accuracy in predicting residential burglary using the grid cell of 800-meter by 800-meter in size as one single location.

This rest of the paper is organized as the followings. The related work is described and compared with our algorithm in Section 2. Our proposed CCRBoost

algorithm is thoroughly discussed in Section 3, including its theoretical analysis. The setting and results of our experiments are presented in Section 4. We then conclude our study in Section 5.

## 2 Related Work

Crime forecasting techniques from feature construction and feature selection point of view, can be categorized into statistic mapping, mathematical modeling, and clustering.

Statistic mapping uses historical statistics of the crime occurred at the same location for forecast[1]. It focuses on seasonality of the crime with the assumption that same type of crime recurs regularly with respect to time, while our approach, in addition to time dimension, also factors in spatial neighborhood and other relevant societal and ecological factors.

In [11], mathematical modeling is used to simulate the formatting of the crime hotspots based on the crime frequency obtained from statistical model of individual criminals. These hotspots are density based and overlapped with each other. Therefore, a suppression process is needed to filter out the true hotspots. Later in [8], Mohler proposed a point-based model that eliminates the suppression step. Using the concept in predicting aftershock, this model simulates how the crime spreads out, like diseases, from the initial background events. The hotspots defined in this model are those locations covering most spread points. The approach suits better for capturing the crime patterns with short life cycles at local level while our algorithm is able to capture the long term cycles at global level using our ensemble patterns. Another caveat of this model fitting approach is that the results can be way off when the incorrect initial points are given. In our framework, the built-in feature selection process can discard irrelevant or misrepresenting patterns when learning the global ensemble pattern.

The clustering approach adapted by Kumar is to define the geographic boundaries of each spatial clusters [7]. With these boundaries, the changing of crime densities in a fixed size cluster is considered as the crime trend of this particular cluster. Our global ensemble spatio-temporal pattern is designed to forecast crime for the whole study area.

## 3 Crime Forecasting Using Spatio-temporal Patterns

### 3.1 The Concept of Spatio-temporal Patterns

Our approach is designed to enhance the utility of the near repeat hypothesis formulated in Social Science [12]. This hypothesis suggests that the same type of crime possibly recurs not only at the same neighboring locations but also at a regular interval of time. In addition, crime incidents are closely related with social behaviors and environmental conditions[6]. This implies that crime tends to have similar trends at locations with similar societal and ecological structures. We hypothesize that crime can be foreseen by investigating the trends of its

correlated crime incidences. All of these three dimensions (location, time, and correlated incidences) are taken into account when we define a spatio-temporal pattern. A spatio-temporal pattern is a local pattern that represents the crime pattern at certain locations during certain time period using correlated crime incidences as the indicators. These indicators are used to represent societal and ecological factors of different locations.

The challenges are that how many local patterns there are during certain time period and at which locations are unknown. Additionally, crime is not evenly distributed throughout a city and there are areas that are more attractive than others to criminals[3]. In order to find the possible local patterns, we use the unsupervised clustering approach without involving geographical features to group those locations with similar indicators in the same time period. This group of locations is considered as the spatial distribution of a possible crime pattern. By varying the configuration of the clustering method, the clusters with different sizes during different periods of time can be generated. A classifier is then trained from each cluster and will be used to represent a local crime pattern.

Our next task is to use the spatio-temporal patterns as features to construct a global level spatio-temporal pattern. This global crime pattern should be capable of detecting crime incidences at every location. Which of these patterns should be selected to form the global crime pattern? Those locally learned spatio-temporal patterns could be redundant or overlapping. And, how can this global pattern be constructed? To resolve these two issues at the same time, we propose a confidence-rate boosting approach. We will first formulate the problem and then discuss our boosting algorithm in detail.

### 3.2 Problem Formulation

We denote one crime indicator, a type of relevant crime event, as  $f_p$ . Different indicators of the same location in the same period of time are used to form a vector, denoted as  $x = [f_1, f_2, \dots, f_P]$ , where  $P$  is the number of correlated incidence types. Each vector  $x$  has one class label  $y$  which tells whether this location is a hotspot. Through the clustering process, the vectors with similar indicators are grouped into one cluster, denoted as  $c$ . A local spatio-temporal pattern, denoted as  $r$ , is defined as:  $r = q(c)$ .  $q()$ , in our case, is a classifier induced from the cluster  $c$  and used to extract the crime pattern. This pattern  $r$  is used to identify whether a vector is a hotspot. We denote  $x \in r$  if  $x$  is recognized as hotspot by pattern  $r$ . Otherwise,  $x \notin r$ .

In reality, a crime pattern might not be represented as one single local pattern because this pattern might shift location-wise or change size over time[9]. In order to truly capture the dynamics of crime patterns, we introduce the ensemble spatio-temporal pattern, denoted as  $R = [r_1 \wedge r_2 \wedge \dots \wedge r_i]$ , which is the conjunction of selected spatio-temporal patterns. Thus, if  $x \in R$ , then it must be true that  $x \in r_1$  and  $x \in r_2$  and  $\dots$  and  $x \in r_i$ . This ensemble pattern is served as the base learner in our confidence-rate boosting approach.

### 3.3 Confidence-Rate Boosting

Before boosting, we balance the training data by setting the weight of hotspots as  $\frac{1}{2H}$  and  $\frac{1}{2C}$  for coldspots,  $H$  and  $C$  are the numbers of hotspots and coldspots, respectively. Each vector  $x_i$  in the training data is denoted as  $D(i)$ . The main idea of our confidence-rate boosting approach is to repeatedly pick the best hypothesis  $h_t$  which yields the least error rate at each boosting round  $t$ . In our case,  $h_t$  is an ensemble spatio-temporal pattern  $R_t$  built at round  $t$ . The error function is defined as:

$$E_{i \sim D_t}[y_i h_t(x_i)] = \sum_i D_t(i) y_i h_t(x_i), \tag{1}$$

where  $D_t$  is the weight distribution at boosting round  $t$ . The theoretical background of the confidence-rate boosting approach is analyzed as follows.

Based on the study in [10], it has been proved that  $\sum_i D_t(i) y_i h_t(x_i) \leq \prod_t Z_t$  so the upper bound of the error rate is  $\prod_t Z_t$ .  $Z_t$  is defined as:

$$Z_t = \sum_i D_t(i) \exp(-\alpha_t y_i h_t(x_i)) \tag{2}$$

Thus, a smaller  $Z_t$  that has a lower error upper bound will lead to a smaller training error at each boosting round. Now, we let  $C_R = \alpha_t h_t(x_i)$  and ignore the boosting round  $t$ . Then, we define our loss function as:

$$Z = \sum_i D(i) \exp(-C_R y_i) \tag{3}$$

and we want to find the minimum value of  $Z$  to lower the training error as much as possible.  $C_R$  is the confidence-rate for pattern  $R$  and  $C_R = 0$  if  $x_i \notin R$ . Here,  $x_i \in R$  means that  $x_i$  is recognized by pattern  $R$  as a hotspot and then set  $y_i = 1$ . Otherwise, set  $y_i = -1$ . Since  $C_R = 0$  where  $x_i \notin R$ , we obtain

$$Z = \sum_{i|x_i \notin R} D(i) + \sum_{i|x_i \in R} D(i) \exp(-C_R y_i) \tag{4}$$

Equation (4) can be rewritten as:

$$Z = W_0 + W_+ \exp(-C_R) + W_- \exp(C_R), \tag{5}$$

where  $W_0 = \sum_{i|x_i \notin R} D(i)$  so  $W_0$  is the total weights of predicted coldspots. And,

$$W_+ = \sum_{i|x_i \in R \text{ and } y=1} D(i), W_- = \sum_{i|x_i \in R \text{ and } y=-1} D(i), \tag{6}$$

$W_+$  is the total weights of true hotspots (true positives), and  $W_-$  is the total weights of false hotspots (false positives). By taking the first derivative of  $Z$  with respect to  $C_R$  and let  $\frac{dZ}{dC_R} = 0$ , we can find the value of  $C_R$  when  $Z$  has the maximum or minimum value:

$$\begin{aligned}
 \frac{dZ}{dC_R} &= -W_+ \exp(-C_R) + W_- \exp(C_R) = 0 \\
 \implies W_- \exp(C_R) &= W_+ \exp(-C_R) \\
 \implies \ln(W_- \exp(C_R)) &= \ln(W_+ \exp(-C_R)) \\
 \implies \ln(W_-) + C_R &= \ln(W_+) - C_R \\
 \implies 2C_R &= \ln(W_+) - \ln(W_-) \\
 \implies C_R &= \frac{1}{2} \ln\left(\frac{W_+}{W_-}\right)
 \end{aligned}$$

And then, we take the second derivative of  $Z$ ,  $\frac{d^2Z}{dC_R^2} = W_+ \exp(-C_R) + W_- \exp(C_R) > 0$ . Since the second derivative of  $Z$  is greater than zero,  $Z$  has the minimum value of  $W_0 + 2\sqrt{W_+W_-}$  when  $C_R = \frac{1}{2} \ln\left(\frac{W_+}{W_-}\right)$ .

To prevent the situation of  $W_- = 0$ , we adjust the above equation as:

$$\hat{C}_R = \frac{1}{2} \ln\left(\frac{W_+ + \frac{1}{2n}}{W_- + \frac{1}{2n}}\right), \quad (7)$$

where  $n$  is the total number of vectors. Equation (7) is then used to calculate the confidence-rate  $\hat{C}_{R_t}$  for pattern  $R_t$  at each round  $t$ .

### 3.4 CCRBoost Algorithm

As described in Algorithm 1, the first task of the CCRBoost algorithm is to identify spatio-temporal patterns of different sizes and shapes with respect to spatial space during each period of time. To add spatio-temporal dimension to our feature, a clustering step is adopted to find the crime distributions at local level in different time periods. K-Means, but not limited to, is chosen to find these patterns. We perform K-Means  $K$  times to obtain  $1+2+\dots+K$  clusters and then train classifiers from each cluster to extract local spatio-temporal patterns at different granularity levels. The data is divided into  $M$  subsets before clustering by certain length of time interval. For example, if the raw crime data is processed by month, then  $M$  equals to 12 when one year worth of data is used. As a result, there are total  $M \times (1 + 2 + \dots + K)$  possible patterns acquired from these  $M$  subsets.

Next, the weights of the entire data set is set to be in a probability distribution which makes the total weight equals to 1. The data set is then randomly divided into two subsets, **GrowSet** and **PruneSet**. This split is based on the total weight instead of the number of records. By calling *BuildChain()*, an ensemble spatio-temporal pattern  $R$  is built from those local patterns. This  $R$  gives the minimum  $Z$  value while evaluating  $R$  on **GrowSet**. Furthermore, *PruneChain()* is called to trim the list of  $R$  and prevents  $R$  from over fitting by using **PruneSet** to reevaluate  $R$  and then obtain the final ensemble pattern  $R_t$ . The confidence-rate  $\hat{C}_{R_t}$  is then calculated by evaluating  $R_t$  on the entire data set using Equation (7). Based on  $\hat{C}_{R_t}$ , the boosting algorithm updated the weights of those vectors that are classified as hotspots. This weight update function is defined as:

$$D_{t+1}(i) = \frac{D_t(i)}{\exp(y_i \hat{C}_{R_t})}, \text{ if } x_i \in R_t \quad (8)$$

The goal is to exponentially lower the weights on those vectors that are recognized by the current global pattern. This way, the data instances which have not been fitted to the pattern are getting more attentions for the next round.

The boosting process is repeated for  $T$  times, where  $T$  is a user-defined variable. At the end,  $T$  ensemble spatio-temporal patterns,  $R_1, R_2, \dots, R_T$ , and  $T$  confidence-rates,  $\hat{C}_{R_1}, \hat{C}_{R_2}, \dots, \hat{C}_{R_T}$  are produced. The formula of the final global spatio-temporal pattern is defined as:

$$h(x) = \begin{cases} \text{hotspot} & \sum_{R_t|x \in R_t} \hat{C}_{R_t} > \alpha \\ \text{coldspot} & \text{otherwise} \end{cases}, \quad (9)$$

while  $\alpha$  is a user-defined threshold.

By taking an input vector  $x$ , this formula evaluates  $x$  over each ensemble pattern  $R_t$ . If  $x$  is recognized by  $R_t$  as a hotspot, then  $\hat{C}_{R_t}$  is added to the total confidence score  $h(x)$ .  $x$  is predicted as a hotspot if  $h(x)$  is greater than the threshold  $\alpha$ . Normally, this threshold  $\alpha$  is set to zero. This ensemble learning algorithm is inspired by Cohen and Singer's research in [2]. The steps of the CCRBoost algorithm are given in Algorithm 1.

---

#### Algorithm 1. CCRBoost Algorithm

---

1. Given crime data  $(x_1, y_1), \dots, (x_n, y_n)$ .
2.  $K$  is a user-defined variable and  $M$  is the total number of time periods.
3. **for**  $k = 1 \dots K$  **do**
4.   **for**  $m = 1 \dots M$  **do**
5.     Run K-Means using the vectors in period  $m$  to generate  $k$  clusters. Then,  $k$  spatio-temporal patterns are extracted from these clusters.
6.   **end for**
7. **end for**
8. Balance the data set by weights.
9. **for**  $t = 1 \dots T$  **do**
10.   Normalize the weights, let  $D_t$  be a probability distribution.
11.   Divide weighted data into two sets, *GrowSet* and *PruneSet*.
12.   Call *BuildChain()* and then *PruneChain()* to obtain  $R_t$ .
13.   Calculate  $\hat{C}_{R_t}$  using entire data set and Equation (7).
14.   Update the weights based on Equation (8).
15. **end for**
16. The final global spatio-temporal pattern is defined as:

$$h(x) = \begin{cases} \text{hotspot} & \sum_{R_t|x \in R_t} \hat{C}_{R_t} > \alpha \\ \text{coldspot} & \text{otherwise} \end{cases}, \quad \alpha \text{ is a user-defined threshold.}$$


---

## 4 Case Study: Forecasting Residential Burglary in a Northeastern City of the U.S.A.

**Data Configuration:** 4-years' (January 2006 to December 2009) crime records have been used for the evaluation. In addition, three different grid resolutions have been applied to generate three data sets from the original crime records.

These three resolutions have the squared cell/block with edge lengths of 800, 600, and 450 meters, respectively.

The targeting crime to be predicted is residential burglary in our experiments. Residential burglary is a particularly interesting crime to study from a prediction perspective since the near repeat hypothesis suggests that proximity to a burgled residence increases the likelihood of victimization of other domiciles in the neighborhood[12].

Based on the criminology theory[6] and after consulting with the domain experts, six categories of incidences are identified having the higher correlation with residential burglary crime than others. These six categories are arrest, commercial burglary, foreclosure, motor vehicle larceny, 911 call, and street robbery. Thus, the aggregations of these six categories from the crime records are used as the crime indicators in our experiments.

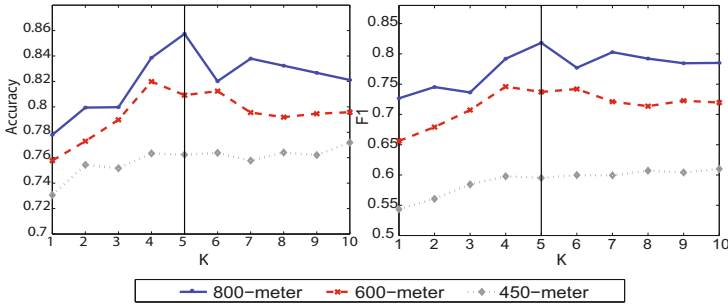
**The Choice of Pattern Learning Classifier:** LADTree[5] has been chosen as the base classifier to identify these patterns in our experiments because LADTree adapts same confidence-rated system to grow a decision tree. However, our algorithm is not limited to LADTree because a spatio-temporal pattern can be represented in any format or model as long as it can tell whether a vector is a hotspot.

**Clustering Approach in Finding Spatio-temporal Patterns:** In this experiment, the effectiveness of  $K$  is evaluated. The other user-defined variables  $T$  (The number of boosting iteration) and  $\alpha$  (The threshold for  $h(x)$ ) are set to 500 and 0, respectively. The results of this experiment are obtained from three data sets with different grid cell sizes. When  $K = 1$ , the data is not clustered. Therefore, the results obtained from the setting of  $K = 1$  is used as the baseline to compare with others. According to the results shown in Figure 2, the clustering approach yields not only the better overall accuracy but also the better F1-score on hotspots. This is because using clustering enhances the feature with spatial dimension by taking into account the crime distribution at local levels. Moreover, we found that the performance reaches certain level when  $K = 4$  and then maintains at this level when  $K \geq 5$ . This shows that the patterns lose the true representative of local crime distributions when the resolution is set too high and suggests that there are less than or equal to 5 different levels of local crime patterns in our target city. Thus,  $K$  is set to 5 in the rest of our experiments.

**Comparing Spatio-Temporal Pattern Features with Random Sampling Features:** In this experiment, the variable  $K$  is used to decide the number of random sampling data sets. This sampling method randomly selects 50% of the data records from a monthly data set for  $1 + 2 + \dots + K$  times without replacement, which means that there is no duplicated records in each sample. This method constructs  $1 + 2 + \dots + K$  samples with unified size and then trains the base classifiers from them using the LADTree algorithm. The purpose is to have same number of features while comparing random sampling with the proposed spatio-temporal pattern.

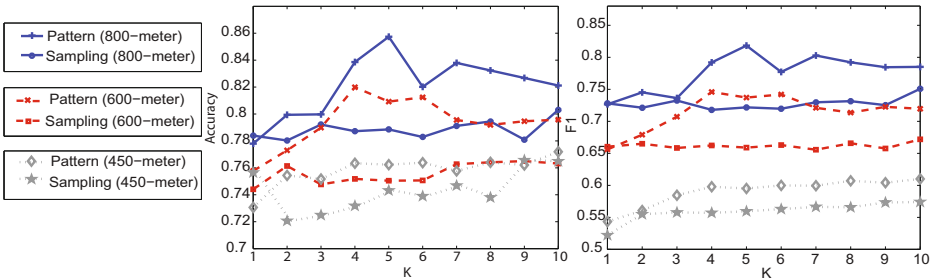
Next, our confidence-rate boosting algorithm is used to pick features from those patterns generated from random sampled clusters and then build a global





**Fig. 2.** The results of using different  $K$  for clustering on 3 data sets

pattern. By this way, we can tell that which kind of feature delivers the better prediction results. According to Figure 3, using spatio-temporal patterns has better performance regardless the resolution of the data set. Thus, spatio-temporal patterns do have the advantage over random sampling due to their spatio-temporal multi-dimensional characteristic.



**Fig. 3.** Comparing spatio-temporal patterns with random sampling on different datasets

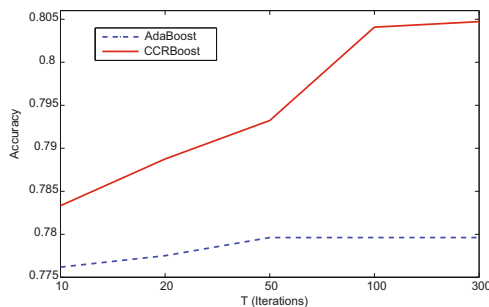
**Comparing with Classification Approaches:** Using the same crime data sets, other commonly used classification methods are adapted to generate the prediction results to compare with our proposed algorithm. Support Vector Machine (SVM), C4.5, Naive Bayes classifier, and LADTree[5] are chosen in this experiment. As shown in Table 1, our proposed CCRBoost algorithm has the best accuracy and F1-score over other classifiers on all three data sets.

**Comparing CCRBoost with AdaBoost:** During this experiment, the various numbers of iterations,  $T$ , are used in comparing our algorithm with the AdaBoost[4] algorithm. LADTree is chosen as the base learning classifier in both algorithms. Shown in Figure 4, the accuracy obtained from the AdaBoost algorithm reaches its ceiling when  $T > 50$ . However, our CCRBoost algorithm not only can obtain better accuracy but also has better convergence rate throughout

**Table 1.** The results of comparing CCRBoost with existing classifiers

Data Set	800-meter		600-meter		450-meter	
Method	Accuracy	F1	Accuracy	F1	Accuracy	F1
SVM	0.817	0.801	0.776	0.742	0.651	0.489
C4.5	0.500	0.667	0.500	0.667	0.500	0
NaiveBayes	0.730	0.675	0.703	0.647	0.667	0.592
LADTree	0.772	0.757	0.728	0.702	0.644	0.487
CCRBoost	<b>0.857</b>	<b>0.818</b>	<b>0.820</b>	<b>0.746</b>	<b>0.772</b>	<b>0.610</b>

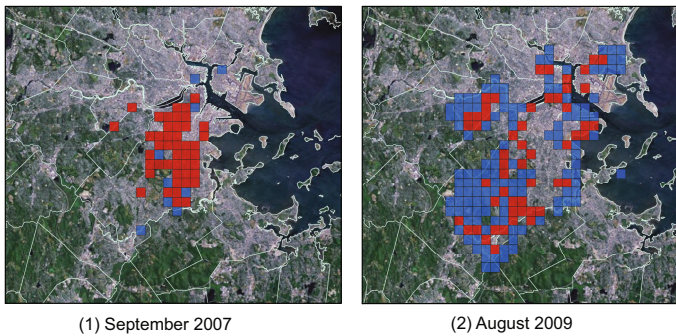
three data sets. In conclusion, the boosting effect of our algorithm is more efficient than AdaBoost because our algorithm enhanced with new spatio-temporal features has a strong impact in predicting crime.



**Fig. 4.** Comparing AdaBoost with CCRBoost on different iterations  $T$

**The Resulting Global Spatio-temporal Pattern:** The selection of the local patterns used in the final spatio-temporal pattern has been visualized on the map, which is shown in Figure 5. The red grid cells represent hotspots and blue cells are coldspots. The first pattern chosen by the proposed algorithm is a cluster from September 2007. The locations of this cluster are consistent with known crime pattern of our target city. The second cluster representing August 2009 data identifies crime hotspots that were excluded from the first cluster. More importantly, this second cluster is useful for pinpointing coldspots areas that have some protective factor against residential burglary and other crimes.

As a result, the first two clusters are complementary in identifying locations where we would expect residential burglary across the entire city as well as areas that are coldspots. Interestingly and consistent with criminological literature, both clusters are in the summer months when children are out of school and individuals may take vacations and be less vigilant about protecting their property. It may be that there is an increased likelihood of residential burglary in this city during the summer time. Based on the consistency with actual crime patterns, our algorithm does find the patterns which recognize not only the spatial but also the temporal factors that are useful for criminal justice professionals in predicting the incidence of future crime.



**Fig. 5.** The first two local patterns used in the final global spatio-temporal pattern resulting from 800-meter grid data set. The red blocks are hotspots and blues are coldspots. (Better viewed in color).

## 5 Conclusions

From a practical standpoint, the patterns selected from this algorithm are indicative of the true locations of residential burglaries throughout the target city. This gives the concrete evidence that using proposed spatio-temporal pattern has the great potential in predicting crime. The ultimate goal of our research is to build a crime prediction system with strong predictive power, which is able to provide forecast in a timely manner and requires less amount of data inputs. Ultimately, the law enforcement is able to fight criminals pro-actively instead of passively.

**Acknowledgments.** The work was partially funded by the National Institute of Justice (No.2009- DE-BX-K219).

## References

1. Cohen, J., Gorr, W.L.: Development of crime forecasting and mapping systems for use by police. H. John Heinz III School of Public Policy and Management. Carnegie Mellon University (2005)
2. Cohen, W., Singer, Y.: A simple, fast, and effective rule learner. In: Proceedings of The Sixteenth National Conference on Artificial Intelligence (1999)
3. Eck, J., Chainey, S., Cameron, J., Wilson, R.: Mapping crime: Understanding hotspots (2005)
4. Freund, Y., Schapire, R.: A decision-theoretic generalization of on-line learning and an application to boosting. In: Computational Learning Theory: Eurocolt, pp. 23–37. Springer (1995)
5. Holmes, G., Pfahringer, B., Kirkby, R., Frank, E., Hall, M.: Multiclass alternating decision trees. In: Elomaa, T., Mannila, H., Toivonen, H. (eds.) ECML 2002. LNCS (LNAI), vol. 2430, pp. 161–172. Springer, Heidelberg (2002)

6. Kelling, G., Coles, C.: *Fixing Broken Windows: Restoring Order and Reducing Crime in Our Communities*. Free Press (1998)
7. Kumar, M.V., Chandrasekar, C.: Spatial clustering simulation on analysis of spatio-temporal crime hotspot for predicting crime activities. *International Journal of Computer Science and Information Technologies* 2(6), 2864–2867 (2011)
8. Mohler, G.O., Short, M.B., Brantingham, P.J., Schoenberg, F.P., Tita, G.E.: Self-exciting point process modeling of crime. *Journal of the American Statistical Association* 106(493) (2011)
9. Ratcliffe, J.H.: Aoristic signatures and the spatio-temporal analysis of high volume crime patterns. *Journal of Quantitative Criminology* 18(1), 23–43 (2002)
10. Schapire, R.E., Singer, Y.: Improved boosting algorithms using confidence-rated predictions. *Machine learning* 37(3), 297–336 (1999)
11. Short, M.B., Bertozzi, A.L., Brantingham, P.J.: Nonlinear patterns in urban crime: Hotspots, bifurcations, and suppression. *SIAM Journal on Applied Dynamical Systems* 9(2), 462–483 (2010)
12. Townsley, M., Homel, R., Chaseling, J.: Infectious burglaries. A test of the near repeat hypothesis. *British Journal of Criminology* 43(3), 615–633 (2003)